Biometrika (2018), **105**, 4, *pp.* 859–872 doi: 10.1093/biomet/asy037
Printed in Great Britain Advance Access publication 14 August 2018

Wild residual bootstrap inference for penalized quantile regression with heteroscedastic errors

By LAN WANG

School of Statistics, University of Minnesota, 224 Church Street South East, Minneapolis, Minnesota 55455, U.S.A.

wangx346@umn.edu

INGRID VAN KEILEGOM

Research Centre for Operations Research and Business Statistics, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium

ingrid.vankeilegom@kuleuven.be

AND ADAM MAIDMAN

School of Statistics, University of Minnesota, 224 Church Street South East, Minneapolis, Minnesota 55455, U.S.A.

maidm004@umn.edu

Summary

We consider a heteroscedastic regression model in which some of the regression coefficients are zero but it is not known which ones. Penalized quantile regression is a useful approach for analysing such data. By allowing different covariates to be relevant for modelling conditional quantile functions at different quantile levels, it provides a more complete picture of the conditional distribution of a response variable than mean regression. Existing work on penalized quantile regression has been mostly focused on point estimation. Although bootstrap procedures have recently been shown to be effective for inference for penalized mean regression, they are not directly applicable to penalized quantile regression with heteroscedastic errors. We prove that a wild residual bootstrap procedure for unpenalized quantile regression is asymptotically valid for approximating the distribution of a penalized quantile regression estimator with an adaptive L_1 penalty and that a modified version can be used to approximate the distribution of a L_1-penalized quantile regression estimator. The new methods do not require estimation of the unknown error density function. We establish consistency, demonstrate finite-sample performance, and illustrate the applications on a real data example.

Some key words: Adaptive lasso; Confidence interval; Lasso; Penalized quantile regression; Wild bootstrap.

1. Introduction

We consider the quantile regression model $Y_i = x_i^{\mathrm{T}}\beta_0 + \epsilon_i$ $(i = 1, \ldots, n)$, where $x_i = (x_{i0}, x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ with $x_{i0} = 1$ is the ith nonstochastic design point in $\mathbb{R}^p$, and ϵ_i is a random error with probability density f_i and the τth quantile equal to zero. The unknown regression coefficient $\beta_0 = (\beta_{00}, \beta_{01}, \ldots, \beta_{0p})^{\mathrm{T}}$ may depend on τ, but we omit such dependence in our notation for simplicity. Quantile regression was proposed by Koenker & Bassett (1978) and has become a popular alternative to least squares regression. Conditional quantiles are of interest in a variety of applications, such as the conditional median of medical expenditure or a low conditional quantile of birth weight. Comparing such quantiles for a range of τ values enables researchers to obtain a more complete picture of the conditional distribution than does mean regression and is particularly useful for analysing heterogeneous data. See Koenker (2005) and Koenker et al. (2017).

We suppose that some of the covariates are irrelevant for modelling the τth conditional quantile but we have no prior information on which. In such a setting, penalized quantile regression has been proven to avoid overfitting by shrinking the estimated coefficients of irrelevant covariates toward zero. Here, we focus on the asymptotic regime where the number of predictors p is fixed while the sample size n goes to infinity. Asymptotic theory for penalized quantile regression in this set-up was recently studied by Zou & Yuan (2008) for independent and identically distributed random errors, and by Wu & Liu (2009), who established the asymptotic distribution of the penalized quantile regression estimator for the adaptive L_1 penalty (Zou, 2006) and considered an extension to the general heteroscedastic error setting. However, these works have not considered estimation of the standard error of the estimated penalized quantile regression coefficients. The asymptotic distribution of L_1-penalized quantile regression has a positive probability mass at zero for the component for which the true regression parameter has a zero value. Inference based directly on asymptotic theory is not convenient. On the other hand, the adaptively L_1-penalized quantile regression estimator enjoys the oracle property under regularity conditions: the zero coefficients are estimated as exactly zero with probability approaching unity, and the nonzero coefficients have the asymptotic normal distribution we would obtain if we knew in advance which coefficients are zero. However, convergence to the oracle distribution is often slow and results in inaccurate confidence intervals (Chatterjee & Lahiri, 2013).

In practice, a two-step procedure is commonly used to construct confidence intervals. First, penalized quantile regression is applied to select variables. Then the model is refitted with the selected variables only to construct confidence intervals. Such a procedure does not account for uncertainties involved in variable selection and generally tends to produce wider confidence intervals, as demonstrated in our simulation study.

These challenges motivate us to develop a wild residual bootstrap-based inference approach for penalized quantile regression with an L_1 or adaptive L_1 penalty. Our work is mostly related to Chatterjee & Lahiri (2010, 2011, 2013) and Camponovo (2015) on bootstrapping penalized estimators in the least squares regression setting. An alternative perturbation method for inference on regularized regression estimates was studied in Minnier et al. (2011). Chatterjee & Lahiri (2010) proved that the standard bootstrap is inconsistent for estimating the distribution of the L_1-penalized least squares estimator when one or more of the components of the regression parameter vector are zero; the failure of the naive paired bootstrap was proved in Camponovo (2015). Modified residual and paired bootstraps were proposed in Chatterjee & Lahiri (2011) and Camponovo (2015), respectively. Chatterjee & Lahiri (2013) demonstrated that although the adaptively penalized least squares estimator enjoys the oracle property, inference based directly on the oracle distribution is often inaccurate, and more accurate inference can be obtained via a residual bootstrap. However, these bootstrap methods do not directly apply to the quantile regression setting due to the nonsmoothness of the quantile loss function and the heteroscedastic error distribution. We prove that a wild residual bootstrap procedure proposed by Feng et al. (2011) for unpenalized quantile regression is asymptotically valid for approximating the distribution of the quantile regression estimator with adaptive L_1 penalty. Furthermore, a modified version of this wild residual bootstrap procedure can be used to approximate the distribution of L_1-penalized quantile regression. Our derivation of the bootstrap consistency theory for penalized quantile regression uses techniques substantially different from that of Feng et al. (2011).

2. Inference for adaptive L_1-penalized quantile regression

2·1. *Quantile regression with adaptive L_1 penalty*

The unpenalized quantile regression estimator for β_0 is $\bar{\beta} = (\bar{\beta}_0, \ldots, \bar{\beta}_p)^{\mathrm{T}}$, where

$$\bar{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau (Y_i - x_i^{\mathrm{T}} \beta) \tag{1}$$

and $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the quantile loss function. Under general regularity conditions, $\bar{\beta}$ is asymptotically normal. The asymptotic covariance matrix of $\bar{\beta}$ depends on the unknown conditional density function of ϵ_i (Koenker, 2005).

Often not all covariates collected are relevant for modelling the τth conditional quantile; that is, some of the components of β_0 are zero. Let $A = \{1 \leqslant j \leqslant p : \beta_{0j} \neq 0\}$ be the index set of the nonzero coefficients. Let $|A| = q$ be the cardinality of the set A. Without loss of generality, we assume that the last $p-q$ components of β_0 are zero; that is, we can write $\beta_0 = (\beta_{01}^{\mathrm{T}}, 0_{p-q}^{\mathrm{T}})^{\mathrm{T}}$, where 0_{p-q} denotes a $(p-q)$- dimensional vector of zeros, and $A = \{1, \ldots, q\}$. Let $X = (x_1, \ldots, x_n)^{\mathrm{T}}$ be the $n \times (p+1)$ matrix of covariates, where $x_1^{\mathrm{T}}, \ldots, x_n^{\mathrm{T}}$ are the rows of X. We also write $X = (1, X_1, \ldots, X_p)$, where $1, X_1, \ldots, X_p$ are the columns of X with 1 representing an n-vector of ones. Define X_A to be the submatrix of X that consists of its first $q+1$ columns; and define X_{A^c} to be the submatrix of X that consists of its last $p-q$ columns. Similarly, let x_{iA} be the subvector that contains the first $q+1$ entries of x_i.

The quantile regression estimator with the adaptive L_1 penalty performs simultaneous estimation and variable selection by minimizing a penalized quantile loss function, i.e.,

$$\tilde{\beta} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \rho_\tau (Y_i - x_i^{\mathrm{T}} \beta) + \lambda_n \sum_{j=1}^{p} w_j |\beta_j| \right\}, \tag{2}$$

where $\lambda_n > 0$ is a tuning parameter and $w_j = |\bar{\beta}_j|^{-\gamma}$ are the adaptive weights ($\gamma > 0$). Write $\tilde{\beta} = (\tilde{\beta}_0, \ldots, \tilde{\beta}_p)^{\mathrm{T}}$ and $\tilde{A} = \{1 \leqslant j \leqslant p : \tilde{\beta}_j \neq 0\}$. Let $\tilde{\beta}_1$ be the subvector that contains the first $q+1$ elements of $\tilde{\beta}$. Let $D_0 = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} x_{iA} x_{iA}^{\mathrm{T}}$ and $D_1 = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} f_i(0) x_{iA} x_{iA}^{\mathrm{T}}$, where $f_i(0)$ is the density function of ϵ_i evaluated at zero. The following properties of $\tilde{\beta}$ were established in Wu & Liu (2009).

Lemma 1. *Assume Condition 2 of §2·2 is satisfied. If $n^{-1/2}\lambda_n \to 0$ and $n^{(\gamma-1)/2}\lambda_n \to \infty$, then the adaptive L_1-penalized quantile regression estimator $\tilde{\beta}$ enjoys the oracle property. That is,*

(i) *$\mathrm{pr}(\tilde{A} = A) \to 1$ as $n \to \infty$;*
(ii) *$n^{1/2}(\tilde{\beta}_1 - \beta_{01}) \to N\{0_{q+1}, \tau(1-\tau) D_1^{-1} D_0 D_1^{-1}\}$ in distribution as $n \to \infty$.*

The result in Lemma 1 is referred to as the oracle property: with probability approaching 1 the zero coefficients of β_0 are identified as zero and the nonzero coefficients are identified as nonzero; and we can estimate the nonzero subvector of β_0 as efficiently as if we know the true model in advance. The proof of Lemma 1 is given in the Supplementary Material.

2·2. *A wild residual bootstrap procedure and its consistency*

We use a wild residual bootstrap procedure to approximate the asymptotic distribution of $\tilde{\beta}$. Our procedure is motivated by the work of Feng et al. (2011) for unpenalized quantile regression. To obtain the wild bootstrap sample, we follow the steps below.

1. We first calculate the residuals from the adaptively penalized quantile regression, $\hat{\epsilon}_i = Y_i - x_i^{\mathrm{T}}\tilde{\beta}$ $(i = 1, \ldots, n)$, where $\tilde{\beta}$ is defined by (2).
2. Let $\epsilon_i^* = r_i|\hat{\epsilon}_i|$, where r_i $(i = 1, \ldots, n)$ are generated as a random sample from a distribution with a cumulative distribution function G satisfying Conditions 3–5 below.
3. We generate the bootstrap sample as $Y_i^* = x_i^{\mathrm{T}}\tilde{\beta} + \epsilon_i^*$ $(i = 1, \ldots, n)$.

Using the bootstrap sample, we recalculate the adaptively penalized quantile regression estimator as

$$\tilde{\beta}^* = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \rho_\tau(Y_i^* - x_i^{\mathrm{T}}\beta) + \lambda_n \sum_{j=1}^{p} w_j^* |\beta_j| \right\},$$

where $w_j^* = |\bar{\beta}_j^*|^{-\gamma}$ and $\bar{\beta}^* = (\bar{\beta}_0^*, \ldots, \bar{\beta}_p^*)^{\mathrm{T}}$ is the ordinary quantile regression estimator recomputed on the bootstrap sample. For $j = 1, \ldots, p$ and $0 < \alpha < 1$, let $d_j^{*(\alpha/2)}$ and $d_j^{*(1-\alpha/2)}$ be the $(\alpha/2)$th and $(1 - \alpha/2)$th quantiles of the bootstrap distribution of $n^{1/2}(\tilde{\beta}_j^* - \tilde{\beta}_j)$, respectively. We can estimate $d_j^{*(\alpha/2)}$ and $d_j^{*(1-\alpha/2)}$ from a large number of bootstrap samples. An asymptotic $100(1 - \alpha)\%$ bootstrap confidence interval for β_{0j}, $j = 1, \ldots, p$, is given by $\big[\tilde{\beta}_j - n^{-1/2}d_j^{*(1-\alpha/2)}, \tilde{\beta}_j - n^{-1/2}d_j^{*(\alpha/2)}\big]$. As in Feng et al. (2011), we work under the following technical conditions.

Condition 1. The true value β_0 is an interior point of a compact set in $\mathbb{R}^p$. The density of ϵ_i, denoted by $f_i(\cdot)$, is Lipschitz continuous and is bounded away from 0 and ∞ in a neighbourhood around 0 for all i.

Condition 2. For some positive-definite matrices B_0 and B_1, $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n x_i x_i^{\mathrm{T}} \to B_0$ and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n f_i(0) x_i x_i^{\mathrm{T}} \to B_1$. Furthermore, $\sum_{i=1}^n \|x_i\|^3 = O(n)$ and $\max_{1\leqslant i\leqslant n}\|x_i\| = O(n^{1/4})$, where $\|\cdot\|$ is the Euclidean norm.

Condition 3. For some strictly positive constants c_1 and c_2, $\sup\{r \in \mathbb{G} : r \leqslant 0\} = -c_1$ and $\inf\{r \in \mathbb{G} : r \geqslant 0\} = c_2$, where $\mathbb{G}$ is the support of the weight distribution G.

Condition 4. The weight distribution G satisfies $\int_0^{+\infty} r^{-1}\,\mathrm{d}G(r) = -\int_{-\infty}^0 r^{-1}\,\mathrm{d}G(r) = 1/2$ and $E_G(|r|) < \infty$, where the expectation is taken under G.

Condition 5. The τth quantile of the distribution G is zero.

Theorem 1 shows that the conditional distribution of $n^{1/2}(\tilde{\beta}^* - \tilde{\beta})$ provides an asymptotically valid approximation to that of $n^{1/2}(\tilde{\beta} - \beta)$. Let $\tilde{A}^* = \{j = 1, \ldots, p : \tilde{\beta}_j^* \neq 0\}$, and let $\tilde{\beta}_1^*$ be the subvector that contains the first $q + 1$ elements of $\tilde{\beta}^*$. Let $r = \{r_1, \ldots, r_n\}$ be the random bootstrap weights and $z = \{z_1, \ldots, z_n\}$ the random sample. By the wild bootstrap mechanism, the distribution of r is independent of that of z. Let pr_z denote the probability under the joint distribution of z, and let $\mathrm{pr}_{r|z}$ denote the probability of r conditional on z.

THEOREM 1. *If Conditions 1–5 and the assumptions of Lemma 1 are satisfied, then* $\mathrm{pr}_{r|z}(\tilde{A}^* = A) = 1 + o_{\mathrm{pr}_z}(1)$. *Furthermore,*

$$\sup_t \left|\mathrm{pr}_{r|z}\{n^{1/2}(\tilde{\beta}_1^* - \tilde{\beta}_1) \leqslant t\} - \mathrm{pr}_z\{n^{1/2}(\tilde{\beta}_1 - \beta_{01}) \leqslant t\}\right| = o_{\mathrm{pr}_z}(1).$$

Remark 1. Conditions 1 and 2 are slightly weaker than the corresponding conditions in Feng et al. (2011). Under Condition 5, conditional on the data, ϵ_i^* has the τth quantile equal to zero. Conditions 3 and 4 ensure that the asymptotic distribution of the bootstrap estimator, conditional on the data, matches the unconditional asymptotic distribution of the original adaptively penalized quantile regression estimator, which depends on the unknown error density function. A simple weight distribution that satisfies Conditions 3–5 is the two-point distribution with probabilities $1-\tau$ and τ at $r = 2(1-\tau)$ and -2τ, respectively. Another example given in Feng et al. (2011) is the distribution such that for $1/8 < \tau < 7/8$, $g(r) = G'(r) = -rI(-2\tau - 1/4 \leqslant r \leqslant -2\tau + 1/4) + rI\{2(1-\tau) - 1/4 \leqslant r \leqslant 2(1-\tau) + 1/4\}$. We propose several other distributions that satisfy these conditions in the Supplementary Material.

Remark 2. By definition $n^{1/2}(\tilde{\beta}^* - \tilde{\beta})$ minimizes $Q_n^*(\delta)$, where $Q_n^*(\delta) = \sum_{i=1}^n \left\{\rho_\tau(\epsilon_i^* - n^{-1/2}x_i^{\mathrm{T}}\delta) - \rho_\tau(\epsilon_i^*)\right\} + \lambda_n \sum_{j=1}^p w_j^*\left(|\tilde{\beta}_j + n^{-1/2}\delta_j| - |\tilde{\beta}_j|\right)$. The crux of the proof of Theorem 1 is to show that conditional on the data,

$$Q_n^*(\delta) \to Q^*(\delta) = \begin{cases} -\delta^T H + \delta^T B_1 \delta/2, & \delta_j = 0 \text{ for } j > q, \\ +\infty, & \text{otherwise} \end{cases}$$

in probability, where $H \sim N\{0, \tau(1-\tau)B_0\}$. Then the results follow from epi-convergence theory; see the 1996 unpublished University of Minnesota technical report of Geyer and the 1999 University of Toronto technical report of Knight.

Remark 3. As pointed out by a referee, Leeb & Pötscher (2008) and Pötscher & Schneider (2009) revealed that the distribution of the adaptive lasso and other shrinkage-type estimators cannot be estimated uniformly in a shrinking neighbourhood of the underlying parameter values. In the setting we consider, the number of covariates is fixed. We assume that the smallest nonzero signal does not diminish to zero as the sample size increases. Furthermore, as in Chatterjee & Lahiri (2011), we do not claim the bootstrap-based estimator of the distribution of the adaptive lasso to be uniformly consistent over any diminishing neighbourhood of the underlying parameter values. See also Remark 3 of Chatterjee & Lahiri (2011).

Remark 4. For the adaptive lasso, the coverage probability of the confidence interval approaches unity, because the wild residual bootstrap distribution approximates the adaptive lasso estimator distribution, which identifies zero coefficients as exactly zero with probability approaching unity.

3. MODIFIED WILD RESIDUAL BOOTSTRAP FOR L_1-PENALIZED QUANTILE REGRESSION

We also consider the L_1- or lasso-penalized quantile regression estimator

$$\check{\beta} = \arg\min_\beta \left\{\sum_{i=1}^n \rho_\tau(Y_i - x_i^{\mathrm{T}}\beta) + \lambda_n \sum_{j=1}^p |\beta_j|\right\},$$

where $\lambda_n > 0$ is a tuning parameter. The asymptotic distribution of $\check{\beta}$ follows that of the minimizer of a random process, which is specified in the following lemma.

LEMMA 2. *Under Condition 2, if $n^{-1/2}\lambda_n \to \lambda_0 \geqslant 0$,*

$$n^{1/2}(\check{\beta} - \beta_0) \to \arg\min_{\delta} \Big[-\delta^T H + \delta^T B_1 \delta/2 + \lambda_0 \sum_{j=1}^{p} \{|\delta_j| I(\beta_{0j} = 0) + \delta_j \operatorname{sign}(\beta_{0j}) I(\beta_{0j} \neq 0)\} \Big]$$

in distribution as $n \to \infty$, where H is defined in Remark 2.

The proof is given in the Supplementary Material. For L_1-penalized mean regression, Chatterjee & Lahiri (2010) proved that the asymptotic distribution of the naive residual bootstrapped lasso estimator is a random measure on $\mathbb{R}^p$ and that the bootstrap is inconsistent whenever the regression parameter vector contains one or more zeros. An explanation of this phenomenon is that the lasso estimates the sign of nonzero coefficients correctly with high probability, but estimates the zero coefficients to be positive or negative with positive probabilities. The naive residual bootstrap fails to reproduce the sign of zero coefficients with high probability. To remedy this, Chatterjee & Lahiri (2010) proposed a thresholding procedure, which we adapt.

Our procedure is as follows. Let $\{a_n\}$ be a sequence of numbers such that $a_n + (n^{-1/2}\log n)a_n^{-1} \to 0$ as $n \to \infty$. For example, $a_n = cn^{-\delta}$ for some $c > 0$ and $0 < \delta < 1/2$. For $\bar{\beta}$ defined in (1), we consider the thresholded estimator $\check{\beta}^* = (\check{\beta}_0^*, \ldots, \check{\beta}_p^*)^{\mathrm{T}}$, where $\check{\beta}_0^* = \bar{\beta}_0$ and $\check{\beta}_j^* = \check{\beta}_j I(|\check{\beta}_j| > a_n)$ for $j = 1 \ldots, p$. Let $\check{\epsilon}_i = Y_i - x_i^{\mathrm{T}}\check{\beta}^*$ $(i = 1, \ldots, n)$. Let $\epsilon_i^{**} = r_i|\check{\epsilon}_i|$ $(i = 1, \ldots, n)$, where the bootstrap weights r_i satisfy Conditions 3–5. We choose to threshold the ordinary quantile regression estimator directly. Alternatively, we may threshold the lasso estimator $\check{\beta}$, which will yield the same asymptotic results for the bootstrapped estimator but requires an additional tuning parameter for the lasso.

The bootstrap sample is generated by $Y_i^{**} = x_i^{\mathrm{T}}\check{\beta}^* + \epsilon_i^{**}$ $(i = 1, \ldots, n)$. We then recalculate the L_1-penalized quantile regression estimator using the bootstrap sample:

$$\check{\beta}^{**} = \arg\min_{\beta} \Big\{ \sum_{i=1}^{n} \rho_\tau(Y_i^{**} - x_i^{\mathrm{T}}\beta) + \lambda_n \sum_{j=1}^{p} |\beta_j| \Big\}.$$

Theorem 2 below shows that the conditional distribution of $n^{1/2}(\check{\beta}^{**} - \check{\beta}^*)$ provides an asymptotically valid approximation to that of $n^{1/2}(\check{\beta} - \beta_0)$.

THEOREM 2. *If Conditions 1–5 and the assumptions of Lemma 2 are satisfied, then*

$$\sup_t \big| \mathrm{pr}_{r|z}\{n^{1/2}(\check{\beta}^{**} - \check{\beta}^*) \leqslant t\} - \mathrm{pr}_z\{n^{1/2}(\check{\beta} - \beta_0) \leqslant t\} \big| = o_{\mathrm{pr}_z}(1).$$

4. NUMERICAL RESULTS

4·1. *Monte Carlo studies*

We study the accuracy of 95% confidence intervals constructed by our bootstrap procedures. For the adaptive L_1 penalty, we select the tuning parameter λ_n by minimizing a Bayesian information criterion (Lee et al., 2014) and consider $\gamma = 1, 2$. For the L_1 penalty, we select λ_n by crossvalidation and consider two choices of a_n. One choice adopts a data-driven approach that minimizes the estimated mean squared error $E^*(\|\check{\beta}^{**} - \check{\beta}^{*}\|^2)$, where E^* is the average over bootstrap samples; see § 5.2 of Chatterjee & Lahiri (2011) and Remark 2 of Camponovo (2015). The other choice is the empirical choice $a_n = n^{-1/3}$, which is motivated by the rate required by the asymptotic theory. The bootstrap random weights r_i are generated from the two-point distribution described in Feng et al. (2011); see Remark 1. We also tried alternative weight distributions and found the results to be similar.

We compare the new methods with the confidence intervals from the oracle model, from the full model, and from the two-step procedure described in § 1 with the adaptive lasso or lasso applied in the first step. The oracle procedure is not implementable in real data analysis. For these competing methods, we consider confidence intervals obtained by the rank-score method and by the wild bootstrap method in the R package quantreg (Koenker, 2016).

Let $Y = 0{\cdot}25X_3 + 0{\cdot}5X_5 + X_7 + 2X_2 + X_1\xi$, where $\xi \sim N(0, 1)$ denotes the random error. Let $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_{10})^{\mathrm{T}} \sim N_{10}(0, I_p)$. We set $X_1 = \Phi(\tilde{X}_1)$, where Φ is the standard normal cumulative distribution function, and $X_i = \tilde{X}_i$ for $i = 2, \ldots, 10$. We consider estimating the conditional median and the $0{\cdot}7$ conditional quantile of Y. The variable X_1 is inactive for estimating the conditional median and is active for estimating the $0{\cdot}7$ conditional quantile. Let $\beta = (\beta_1, \ldots, \beta_{10})^{\mathrm{T}}$ be the vector of quantile regression coefficients. We have $\beta_3 = 0{\cdot}25$, $\beta_5 = 0{\cdot}5$, $\beta_7 = 1$, $\beta_9 = 2$, and $\beta_2 = \beta_4 = \beta_6 = \beta_8 = \beta_{10} = 0$ for both quantiles, and we have $\beta_1 = 0$ for the conditional median and $\beta_1 = \Phi^{-1}(0{\cdot}7)$ for the $0{\cdot}7$ conditional quantile.

We perform 1000 simulations with 400 bootstrapped samples for each. We use sample size $n = 100$ for estimating the conditional median and size 250 for estimating the $0{\cdot}7$ conditional quantile, as it is known to be more challenging to estimate a higher quantile than to estimate the median. Table 1 summarizes the simulation results. The standard errors of the coverage probabilities are below $0{\cdot}01$ and the standard errors of the confidence interval lengths are below $0{\cdot}005$ for all cases. We also report the average number of nonzero coefficients correctly identified to be nonzero and the average number of zero coefficients incorrectly identified to be nonzero. For the two-step procedure, we only report results for $\gamma = 1$ if the adaptive lasso is applied in step 1 as the results for $\gamma = 2$ are similar. Additional simulation results are given in the Supplementary Material.

The wild residual bootstrap procedures achieve the specified coverage probability. For the L_1 penalty, the two choices of a_n yield similar results. The adaptive L_1 penalty produces sparser models than the L_1 penalty. The resulting confidence intervals are generally shorter than those based on the full model or the two-step procedure. For the adaptive lasso, the coverage probability of the confidence interval for zero coefficients is close to 1; see Remark 4. Similar numerical findings for adaptive lasso-penalized least squares regression were reported in Minnier et al. (2011) and Camponovo (2015).

4·2. *Real data example*

We analyse data on the effects of ozone on school children's lung growth (Ihorst et al., 2004). The study was carried out from February 1996 to October 1999 in Germany on school children initially in first and second primary school classes. The data we analyse contain a subset of 496 children with complete data at three examinations (Buchholz et al., 2008).

The response variable is the forced vital capacity of the lung. We consider the ten explanatory variables with the largest inclusion probabilities using the bootstrap procedure from De Bin et al. (2015): gender, x_1; height at pulmonary function testing, x_2; weight at pulmonary function testing, x_3; maximal nitrogen oxide value of last 24 hours before pulmonary function testing, x_4;

Table 1. *Empirical coverage probabilities* ($\times 100$) *and average interval lengths (in parentheses) for nominal* 95% *confidence intervals*

	$\beta_1 = \Phi^{-1}(\tau)$	$\beta_3 = 0{\cdot}25$	$\beta_5 = 0{\cdot}5$	$\beta_7 = 1$	$\beta_9 = 2$	Zeros	TP	FP
			$\tau = 0{\cdot}5$	$n = 100$				
New AL1	92·0 (0·33)	94·6 (0·15)	93·2 (0·17)	95·3 (0·13)	92·7 (0·14)	97·4 (0·06)	4	0·3
New AL2	90·6 (0·42)	95·0 (0·15)	93·6 (0·17)	95·1 (0·13)	92·5 (0·14)	98·3 (0·06)	4	0·3
New L1	90·7 (0·28)	92·9 (0·15)	92·4 (0·18)	94·9 (0·15)	91·2 (0·16)	93·5 (0·11)	4	3·3
New L2	92·2 (0·29)	93·7 (0·16)	93·6 (0·19)	96·1 (0·16)	94·5 (0·17)	95·5 (0·12)	4	3·3
Full RS	94·8 (0·59)	95·9 (0·21)	96·7 (0·24)	96·2 (0·21)	96·1 (0·22)	95·9 (0·21)	4	6
Full WB	91·0 (0·54)	97·4 (0·18)	95·9 (0·22)	97·6 (0·18)	94·6 (0·20)	96·1 (0·19)	4	6
TS AL RS	94·8 (0·51)	96·6 (0·21)	96·3 (0·27)	97·1 (0·23)	95·6 (0·23)	98·2 (0·26)	4	0·3
TS AL WB	91·5 (0·47)	95·5 (0·16)	94·2 (0·21)	96·0 (0·17)	92·4 (0·19)	97·7 (0·21)	4	0·3
TS L RS	94·1 (0·52)	96·2 (0·22)	95·6 (0·27)	96·0 (0·23)	95·4 (0·24)	96·3 (0·26)	4	3·3
TS L WB	92·1 (0·49)	94·7 (0·18)	94·3 (0·22)	95·9 (0·19)	93·3 (0·20)	95·8 (0·21)	4	3·3
Oracle RS	-	97·1 (0·21)	97·9 (0·26)	97·0 (0·20)	97·2 (0·18)	-	4	0
Oracle WB	-	97·7 (0·15)	95·9 (0·19)	98·2 (0·15)	97·2 (0·16)	-	4	0
			$\tau = 0{\cdot}7$	$n = 250$				
New AL1	89·6 (0·35)	94·8 (0·10)	92·2 (0·09)	94·9 (0·08)	93·6 (0·09)	98·7 (0·04)	5	0·1
New AL2	89·8 (0·34)	94·1 (0·09)	91·7 (0·09)	95·0 (0·08)	93·1 (0·09)	99·0 (0·04)	5	0·1
New L1	90·1 (0·34)	94·4 (0·10)	94·2 (0·10)	95·4 (0·08)	95·1 (0·09)	95·4 (0·06)	5	2·6
New L2	90·7 (0·35)	94·9 (0·10)	94·2 (0·10)	95·4 (0·08)	95·1 (0·09)	95·9 (0·06)	5	2·6
Full RS	94·9 (0·39)	96·8 (0·12)	95·3 (0·12)	95·8 (0·10)	96·4 (0·11)	95·9 (0·11)	5	5
Full WB	90·6 (0·37)	96·3 (0·11)	95·5 (0·11)	97·3 (0·09)	96·1 (0·11)	96·2 (0·10)	5	5
TS AL RS	93·8 (0·37)	95·4 (0·12)	96·1 (0·10)	95·9 (0·11)	96·4 (0·12)	98·8 (0·11)	5	0·1
TS AL WB	91·7 (0·35)	95·2 (0·11)	95·7 (0·09)	95·8 (0·10)	96·5 (0·11)	98·9 (0·11)	5	0·1
TS L RS	93·8 (0·37)	95·0 (0·12)	95·3 (0·11)	96·2 (0·11)	95·5 (0·12)	96·1 (0·11)	5	2·6
TS L WB	91·2 (0·35)	94·8 (0·12)	95·2 (0·10)	95·7 (0·11)	96·8 (0·12)	96·0 (0·10)	5	2·6
Oracle RS	94·0 (0·38)	96·8 (0·11)	95·3 (0·11)	95·9 (0·09)	96·4 (0·10)	-	5	0
Oracle WB	90·8 (0·36)	95·7 (0·10)	94·9 (0·10)	96·6 (0·08)	96·4 (0·10)	-	5	0

New AL1, proposed method with adaptive L_1 penalty ($\gamma = 1$); New AL2, proposed method with adaptive L_1 penalty ($\gamma = 2$); New L1, proposed method with L_1 penalty data-driven choice of a_n; New L2, proposed method with L_1 penalty $a_n = n^{-1/3}$; Full RS, full model with rank-score method; Full WB, full model with wild residual bootstrap; TS AL RS, two-step procedure, adaptive L_1 ($\gamma = 1$) followed by rank-score method; TS AL WB, two-step procedure, adaptive L_1 ($\gamma = 1$) followed by wild residual bootstrap; TS L RS, two-step procedure, lasso followed by rank-score method; TS L WB, two-step procedure, lasso followed by wild residual bootstrap; Oracle RS, oracle model with rank-score method; Oracle WB: oracle model with wild residual bootstrap; Zeros, the reported average coverage probability length, the average for all zero coefficients; TP, average number of true positives; FP, average number of false positives.

wheezing or whistling in the chest, x_5; shortness of breath, x_6; whether patient lives in a village with high ozone values, x_7; sensitization to pollen, x_8; sensitization to dust mite allergens, x_9; and age at 1 March 1996, x_{10}.

Table 2 reports 95% confidence intervals for each covariate from bootstrapping penalized quantile regression with the adaptive L_1 and L_1 penalties for estimating the conditional median and the conditional 0·7 quantile. For both methods, the variables x_1, x_2 and x_3 are identified as significant at both quantiles.

Acknowledgement

Wang and Maidman acknowledge financial support from the U.S. National Science Foundation. Van Keilegom was supported by the European Research Council and the Interuniversity Attraction

Table 2. *Analysis of ozone data: wild residual-based* 95% *bootstrapped confidence intervals for the* 0·5 *and* 0·7 *conditional quantiles*

	τ =0·5			τ =0·7		
	New AL1	New AL2	New L	New AL1	New AL2	New L
Intercept	(2·26, 2·31)	(2·26, 2·30)	(2·26, 2·31)	(2·37, 2·41)	(2·37, 2·41)	(2·37, 2·42)
x_1	(−0·13, −0·08)	(−0·12, −0·09)	(−0·10, −0·10)	(−0·12, −0·08)	(−0·12, −0·08)	(−0·10, −0·10)
x_2	(0·15, 0·22)	(0·14, 0·20)	(0·18, 0·24)	(0·16, 0·22)	(0·16, 0·22)	(0·21, 0·26)
x_3	(0·04, 0·12)	(0·05, 0·12)	(0·07, 0·08)	(0·06, 0·15)	(0·06, 0·15)	(0·08, 0·09)
x_4	(0, 0)	(−0·01, 0·01)	(0, 0)	(−0·01, 0)	(−0·01, 0)	(0, 0)
x_5	(0, 0)	(0·01, 0·03)	(0·02, 0·02)	(−0·01, 0)	(−0·01, 0)	(0, 0)
x_6	(0, 0)	(0, 0)	(0, 0)	(0·01, 0·05)	(0·01, 0·05)	(0·03, 0·03)
x_7	(0, 0)	(−0·01, 0·01)	(0, 0)	(0, 0·01)	(−0·01, 0·01)	(0, 0)
x_8	(0, 0)	(−0·01, 0·01)	(0, 0)	(−0·03, −0·01)	(−0·03, 0)	(−0·02, −0·02)
x_9	(0, 0)	(−0·01, 0·01)	(0, 0)	(0, 0·02)	(0, 0·02)	(0, 0)
x_{10}	(0, 0)	(0, 0·04)	(0·01, 0·02)	(0, 0·01)	(0, 0·01)	(−0·01, 0)

New AL1, proposed method with adaptive L_1 penalty ($\gamma = 1$); New AL2, proposed method with adaptive L_1 penalty ($\gamma = 2$); New L, proposed method with L_1 penalty (data-driven choice of a_n).

Pole research network of the Belgian government. The authors are grateful to the editor, associate editor, and referees for their insightful comments.

Supplementary material

Supplementary material available at *Biometrika* online includes proofs, additional examples of random weight distributions and simulation results.

Appendix

Proofs of Theorems 1 and 2

We use E^* and var* to denote expectation and variance conditional on the sample z. Let $E_{r,z}$ and var$_{r,z}$ be the expectation and variance with respect to the joint distribution of r and z. Let pr denote the probability under the joint distribution, and let $\mathrm{pr}_{r|z}$ denote the probability of r conditional on z. A random variable R_n is said to be $o^*_{pr}(1)$ if for any $\epsilon, \delta > 0$, $\mathrm{pr}_z\{\mathrm{pr}_{r|z}(|R_n| > \epsilon) > \delta\} \to 0$ as $n \to \infty$, and $o_{\mathrm{pr}_{r,z}}(1)$ is the regular notion with respect to the joint distribution of r and z. Lemma 3 from Cheng & Huang (2010) will be used repeatedly to allow for the transition of various stochastic orders in different probability spaces.

Let $V^*_n(\delta) = \sum_{i=1}^n \left\{\rho_\tau(\epsilon^*_i - n^{-1/2}x_i^{\mathrm{T}}\delta) - \rho_\tau(\epsilon^*_i)\right\}$. Let $\psi_\tau(u) = \tau - I(u < 0)$. It follows from Knight (1998) and Koenker (2005) that

$$V^*_n(\delta) = -n^{-1/2}\sum_{i=1}^n x_i^{\mathrm{T}}\delta\psi_\tau(\epsilon^*_i) + \sum_{i=1}^n \int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \left\{I(\epsilon^*_i \leqslant s) - I(\epsilon^*_i \leqslant 0)\right\} \mathrm{d}s = V^*_{1n}(\delta) + V^*_{2n}(\delta).$$

Lemma A1. *Under the conditions of Theorem 1,*

$$\sup_t \left|\mathrm{pr}_{r|z}\{V^*_{1n}(\delta) \leqslant t\} - \mathrm{pr}_z\{-\delta^T H \leqslant t\}\right| = o_{\mathrm{pr}_z}(1). \tag{A1}$$

The proof of Lemma A1 is given in the Supplementary Material.

LEMMA A2. *Under the conditions of Theorem 1,*

$$V_{2n}^*(\delta) = \delta^T B_1 \delta/2 + o_{pr}^*(1). \tag{A2}$$

Proof. Recall that $\epsilon_i^* = r_i|\hat{\epsilon}_i|$ and $\hat{\epsilon}_i = \epsilon_i - x_i^{\mathrm{T}}(\tilde{\beta} - \beta_0)$. We will show that

$$\sup_{b \in B} |V_{2n}^*(\delta, b) - \delta^T B_1 \delta/2| = o_{pr}^*(1),$$

where $V_{2n}^*(\delta, b) = \sum_{i=1}^n \int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \left\{I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b| \leqslant s) - I(r_i \leqslant 0)\right\} \mathrm{d}s$, with B a compact set and $\eta > 0$. Since $\mathrm{pr}_z\{n^{1/2-\eta}(\tilde{\beta} - \beta_0) \in B\} \to 1$, the result of the lemma follows. By Lemma 3 of Cheng & Huang (2010), it suffices to show that

$$\sup_{b \in B} |V_{2n}^*(\delta, b) - \delta^T B_1 \delta/2| = o_{pr,z}(1).$$

We will use Theorem 2.11.9 in van der Vaart & Wellner (1996). For a fixed $\varepsilon > 0$, divide the set B in to $O(\varepsilon^{-2p})$ cubes of the form $C_k = \prod_{j=1}^p [b_{j,k_j-1}, b_{j,k_j})$ where $k = (k_1, \ldots, k_p)^{\mathrm{T}}$, $k_j = 1, \ldots, O(\varepsilon^{-2})$ for $j = 1, \ldots, p$, and $b_{j,k_j} - b_{j,k_j-1} \leqslant \varepsilon^2$. Then, writing $V_{2n}^*(\delta, b) = \sum_{i=1}^n v_{ib}$, we will show that

$$\sum_{i=1}^n E_{r,z}\Big(\sup_{b,b' \in C_k} |v_{ib} - v_{ib'}|^2\Big) \leqslant \varepsilon^2. \tag{A3}$$

Indeed, for fixed i and for $b, b' \in C_k$, $|v_{ib} - v_{ib'}|^2$ is bounded above by

$$\begin{aligned}
&\Big|\int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \left\{I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b| \leqslant s) - I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b'| \leqslant s)\right\}\mathrm{d}s\Big|^2 \\
&\quad \leqslant I(x_i^{\mathrm{T}}\delta > 0) n^{-1/2}x_i^{\mathrm{T}}\delta \int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \left|I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b| \leqslant s) - I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b'| \leqslant s)\right|\mathrm{d}s \\
&\qquad + I(x_i^{\mathrm{T}}\delta \leqslant 0) n^{-1/2}|x_i^{\mathrm{T}}\delta| \int_0^{n^{-1/2}|x_i^{\mathrm{T}}\delta|} \left|I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b| \leqslant -s) - I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b'| \leqslant -s)\right|\mathrm{d}s.
\end{aligned}$$

Let us focus on the first term above, as the second term is similar. The first term equals

$$\begin{aligned}
&I(x_i^{\mathrm{T}}\delta > 0, r_i > 0) n^{-1/2}x_i^{\mathrm{T}}\delta \int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \Big|I(-s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b \leqslant \epsilon_i \leqslant s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b) \\
&\qquad\qquad - I(-s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b' \leqslant \epsilon_i \leqslant s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b')\Big|\,\mathrm{d}s \\
&\quad \leqslant I(x_i^{\mathrm{T}}\delta > 0, r_i > 0) n^{-1/2}x_i^{\mathrm{T}}\delta \int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \Big\{\Big|I(\epsilon_i \leqslant s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b) - I(\epsilon_i \leqslant s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b')\Big| \\
&\qquad\qquad + \Big|I(\epsilon_i \leqslant -s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b) - I(\epsilon_i \leqslant -s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b')\Big|\Big\}\,\mathrm{d}s \\
&\quad \leqslant I(x_i^{\mathrm{T}}\delta > 0, r_i > 0) n^{-1/2}x_i^{\mathrm{T}}\delta \int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \Big[\Big\{I(\epsilon_i \leqslant s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b_k) - I(\epsilon_i \leqslant s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b_{k-1})\Big\} \\
&\qquad\qquad + \Big\{I(\epsilon_i \leqslant -s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b_k) - I(\epsilon_i \leqslant -s/r_i + n^{-1/2+\eta}x_i^{\mathrm{T}}b_{k-1})\Big\}\Big]\,\mathrm{d}s,
\end{aligned}$$

where for notational simplicity we assume that all components of x_i are positive. Hence,

$$\begin{aligned}
&\sum_{i=1}^{n} E_{r,z}\Big(\sup_{b,b'\in C_k} |v_{ib}-v_{ib'}|^2\Big)\\
&\quad\leqslant n^{-1/2}\sum_{i=1}^{n}|x_i^{\mathrm T}\delta|\int\int_0^{n^{-1/2}|x_i^{\mathrm T}\delta|}\Big[\big\{F_i(s/r+n^{-1/2+\eta}x_i^{\mathrm T}b_k)-F_i(s/r+n^{-1/2+\eta}x_i^{\mathrm T}b_{k-1})\big\}\\
&\qquad+\big\{F_i(-s/r+n^{-1/2+\eta}x_i^{\mathrm T}b_k)-F_i(-s/r+n^{-1/2+\eta}x_i^{\mathrm T}b_{k-1})\big\}\Big]\,\mathrm{d}s\,\mathrm{d}G(r)\\
&\quad\leqslant 2n^{-1}\sum_{i=1}^{n}|x_i^{\mathrm T}\delta|^2 n^{-1/2+\eta}x_i^{\mathrm T}|b_k-b_{k-1}|\sup_{t\in\mathcal N_i}f_i(t)\leqslant c\varepsilon^2
\end{aligned}$$

for some $0<c<\infty$ and for $\eta\leqslant 1/2$, where $\mathcal N_i$ is a neighbourhood of 0 such that $\sup_{t\in\mathcal N_i}f_i(t)<\infty$; see Condition 1. This verifies (A3).

Let $N_{[\,]}(\varepsilon,B,L_2^n)$ be the bracketing number of B, i.e., the minimal number of sets N_ε in a partition $B=\bigcup_{j=1}^{N_\varepsilon}B_{\varepsilon j}$ such that $\sum_{i=1}^n E_{r,z}\big\{\sup_{b,b'\in B_{\varepsilon j}}(v_{ib}-v_{ib'})^2\big\}\leqslant\varepsilon^2$ for $j=1,\ldots,N_\varepsilon$. For any $\delta_n\downarrow 0$,

$$\int_0^{\delta_n}\{\log N_{[\,]}(\varepsilon,B,L_2^n)\}^{1/2}\,\mathrm{d}\varepsilon\leqslant c\int_0^{\delta_n}\{\log(\varepsilon^{-2p})\}^{1/2}\,\mathrm{d}\varepsilon\to 0.$$

Since the partition of B does not depend on n and since $\sup_{b\in B}|v_{ib}|\to 0$ for all i, it follows from Theorem 2.11.9 in van der Vaart & Wellner (1996) that $V_{2n}^*(\delta,b)-E_{r,z}\{V_{2n}^*(\delta,b)\}$ converges weakly in $\ell^\infty(B)$ provided it converges marginally, where $\ell^\infty(B)$ is the space of bounded functions from B to $\mathbb R$ equipped with the supremum norm.

To check convergence of $V_{2n}^*(\delta,b)$ for fixed $b\in B$, it suffices to show that $E_{r,z}\{V_{2n}^*(\delta,b)\}\to\delta^T B_1\delta/2$ and $\mathrm{var}_{r,z}\{V_{2n}^*(\delta,b)\}\to 0$. Note that

$$\begin{aligned}
&E_{r,z}\{V_{2n}^*(\delta,b)\}\\
&\quad=E_r\Big(E_{z|r}\Big[\sum_{i=1}^n\int_0^{n^{-1/2}x_i^{\mathrm T}\delta}\big\{I(r_i|\epsilon_i-n^{-1/2+\eta}x_i^{\mathrm T}b|\leqslant s)-I(r_i\leqslant 0)\big\}\,\mathrm{d}s\Big]\Big)\\
&\quad=\int_0^\infty\sum_{i=1}^n\int_0^{n^{-1/2}x_i^{\mathrm T}\delta}\big\{F_i(s/r+n^{-1/2+\eta}x_i^{\mathrm T}b)-F_i(-s/r+n^{-1/2+\eta}x_i^{\mathrm T}b)\big\}I(x_i^{\mathrm T}\delta>0)\,\mathrm{d}s\,\mathrm{d}G(r)\\
&\qquad+\int_{-\infty}^0\sum_{i=1}^n\int_0^{n^{-1/2}x_i^{\mathrm T}\delta}\big\{1-F_i(s/r+n^{-1/2+\eta}x_i^{\mathrm T}b)+F_i(-s/r+n^{-1/2+\eta}x_i^{\mathrm T}b)-1\big\}I(x_i^{\mathrm T}\delta<0)\,\mathrm{d}s\,\mathrm{d}G(r)\\
&\quad=W_1+W_2,
\end{aligned}$$

say, where F_i denotes the distribution of ϵ_i. Then

$$\begin{aligned}
W_1&=\int_0^\infty\sum_{i=1}^n\int_0^{n^{-1/2}x_i^{\mathrm T}\delta}\big\{f_i(0)2s/r\big\}I(x_i^{\mathrm T}\delta>0)\,\mathrm{d}s\,\mathrm{d}G(r)\\
&\quad+\int_0^\infty\sum_{i=1}^n\int_0^{n^{-1/2}x_i^{\mathrm T}\delta}\big\{f_i(t^*/r)-f_i(0)\big\}(2s/r)I(x_i^{\mathrm T}\delta>0)\,\mathrm{d}s\,\mathrm{d}G(r)=W_{11}+W_{12},
\end{aligned}$$

say, where t^* is between $-n^{-1/2}x_i^{\mathrm T}\delta+n^{-1/2+\eta}x_i^{\mathrm T}b$ and $n^{-1/2}x_i^{\mathrm T}\delta+n^{-1/2+\eta}x_i^{\mathrm T}b$. Note that

$$W_{11}=\int_0^\infty r^{-1}\mathrm{d}G(r)\sum_{i=1}^n f_i(0)\big(n^{-1/2}x_i^{\mathrm T}\delta\big)^2 I(x_i^{\mathrm T}\delta>0)=\frac12\delta^{\mathrm T}\Big\{n^{-1}\sum_{i=1}^n f_i(0)x_ix_i^T I(x_i^{\mathrm T}\delta>0)\Big\}\delta.$$

By Condition 1, there exists a positive constant c such that

$$
\begin{aligned}
|W_{12}| &\leqslant c \int_0^\infty \sum_{i=1}^n \int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \big(n^{-1/2}x_i^{\mathrm{T}}\delta/r + n^{-1/2+\eta}|x_i^{\mathrm{T}}b|\big) 2s/r I(x_i^{\mathrm{T}}\delta > 0)\,\mathrm{d}s\,\mathrm{d}G(r) \\
&\leqslant c\Big\{\int_0^\infty r^{-2}\,\mathrm{d}G(r)\Big\}\big(n^{-1/2}\|\delta\| \max_{1\leqslant i\leqslant n}\|x_i\|\big)\Big[\delta^{\mathrm{T}}\Big\{n^{-1}\sum_{i=1}^n x_i x_i^T I(x_i^{\mathrm{T}}\delta > 0)\Big\}\delta\Big] \\
&\quad + c\Big\{\int_0^\infty r^{-1}\mathrm{d}G(r)\Big\}\big(n^{-1/2+\eta}\|b\| \max_{1\leqslant i\leqslant n}\|x_i\|\big)\Big[\delta^{\mathrm{T}}\Big\{n^{-1}\sum_{i=1}^n x_i x_i^T I(x_i^{\mathrm{T}}\delta > 0)\Big\}\delta\Big] \to 0,
\end{aligned}
$$

as Conditions 3 and 4 imply that $\int_0^\infty r^{-2}\,\mathrm{d}G(r)$ is bounded, and by Condition 2 we have $n^{-1/2+\eta}\max_{1\leqslant i\leqslant n}\|x_i\| \to 0$ for η small enough. Similarly, we can show that $W_2 = \frac{1}{2}\delta^{\mathrm{T}}\big\{n^{-1}\sum_{i=1}^n f_i(0)x_i x_i^T I(x_i^{\mathrm{T}}\delta < 0)\big\}\delta + o(1)$. Hence, $E_{r,z}\{V_{2n}^*(\delta,b)\} \to \delta^T B_1\delta/2$ as $n \to \infty$. To show that $\mathrm{var}_{r,z}\{V_{2n}^*(\delta,b)\} \to 0$, we observe that

$$
\begin{aligned}
\mathrm{var}_{r,z}\{V_{2n}^*(\delta,b)\} &= \sum_{i=1}^n \mathrm{var}_{r,z}\Big[\int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \big\{I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b| \leqslant s) - I(r_i \leqslant 0)\big\}\,\mathrm{d}s\Big] \\
&\leqslant \sum_{i=1}^n E_{r,z}\Big[\int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \big\{I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b| \leqslant s) - I(r_i \leqslant 0)\big\}\,\mathrm{d}s\Big]^2 \\
&= \Big(n^{-1/2}\|\delta\| \max_{1\leqslant i\leqslant n}\|x_i\|\Big) E_{r,z}\{V_{2n}^*(\delta,b)\},
\end{aligned}
$$

where the last equality follows because $\int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta} \big\{I(r_i|\epsilon_i - n^{-1/2+\eta}x_i^{\mathrm{T}}b| \leqslant s) - I(r_i \leqslant 0)\big\}\,\mathrm{d}s$ is always nonnegative. Since $n^{-1/2}\max_{1\leqslant n}\|x_i\| \to 0$ and $E_{r,z}\{V_{2n}^*(\delta,b)\} \to \delta^T B_1\delta/2$, we have $\mathrm{var}_{r,z}\{V_{2n}^*(\delta,b)\} \to 0$ as $n \to \infty$. This completes the proof. □

Proof of Theorem 1. Recall that $Q_n^*(\delta) = \sum_{i=1}^n \big\{\rho_\tau(\epsilon_i^* - n^{-1/2}x_i^{\mathrm{T}}\delta) - \rho_\tau(\epsilon_i^*)\big\} + \lambda_n \sum_{j=1}^p w_j^*\big(|\tilde\beta_j + n^{1/2}\delta_j| - |\tilde\beta_j|\big)$, where $w_j^* = |\bar\beta_j^*|^{-\gamma}$, $\bar\beta^* = (\bar\beta_0^*, \bar\beta_1^*, \ldots, \bar\beta_p^*)^{\mathrm{T}}$ is the ordinary quantile regression estimator computed from the bootstrap sample, and $\gamma > 0$. We have $n^{1/2}(\tilde\beta^* - \tilde\beta) = \arg\min_\delta Q_n^*(\delta)$. Let A_n denote the event that the adaptive lasso estimator $\tilde\beta$ correctly estimated all the zero components of β, i.e., A_n is the set of all $\omega \in \Omega$ such that $\{j : 1 \leqslant j \leqslant p, \tilde\beta_j(\omega) = 0\} = \{q+1, \ldots, p\}$. Then it follows from Lemma 1 that $\mathrm{pr}(A_n) \to 1$ as $n \to \infty$. There exists a subsequence $\{n_k\}$ such that $\mathrm{pr}(A_{n_k}^c \text{ infinitely often}) = 0$. Let Ω_0^c be the union of $\limsup_k A_{n_k}^c$ and the event on which (A1) or (A2) fails to hold; then $\mathrm{pr}(\Omega_0) = 1$. For any fixed $w \in \Omega_0$, there exists $n_w \geqslant 1$ such that for all $n \geqslant n_w$, $\{j : 1 \leqslant j \leqslant p, \tilde\beta_{nj}(\omega) = 0\} = \{q+1, \ldots, p\}$. Hence on Ω_0, as $n \to \infty$,

$$
Q_n^*(\delta) \to Q^*(\delta) = \begin{cases} -\delta^T H + \delta^T B_1\delta/2, & \delta_{q+1} = \cdots = \delta_p = 0, \\ +\infty, & \text{otherwise} \end{cases}
$$

in probability. Following the same argument as in Lemma 1 and applying epi-convergence theory, as in the unpublished technical reports of Geyer and Knight, the result is established by the equivalent representation of bootstrap consistency in (23.2) of van der Vaart (1998). □

Proof of Theorem 2. Let $A_n = \{\|\breve\beta^* - \beta_0\| \leqslant cn^{-1/2}\log(n)\}$ for some given positive constant c. Since $\bar\beta$ is $n^{1/2}$-consistent, we have $\mathrm{pr}(A_n) \to 1$. Let $Q_n^{**}(\delta) = \sum_{i=1}^n \big\{\rho_\tau(\epsilon_i^{**} - n^{-1/2}x_i^{\mathrm{T}}\delta) - \rho_\tau(\epsilon_i^{**})\big\} + \lambda_n \sum_{j=1}^p \big(|\breve\beta_j^* + n^{-1/2}\delta_j| - |\breve\beta_j^*|\big)$; then $n^{1/2}(\breve\beta^{**} - \breve\beta^*)$ minimizes $Q_n^{**}(\delta)$. Let $V_n^{**}(\delta) = \sum_{i=1}^n \big\{\rho_\tau(\epsilon_i^{**} - x_i^{\mathrm{T}}\delta/n^{1/2}) - \rho_\tau(\epsilon_i^{**})\big\}$.

We can write

$$V_n^{**}(\delta) = -n^{-1/2}\sum_{i=1}^{n} x_i^{\mathrm{T}}\delta\psi_\tau(\epsilon_i^{**}) + \sum_{i=1}^{n}\int_0^{n^{-1/2}x_i^{\mathrm{T}}\delta}\left\{I(\epsilon_i^{**}\leqslant s) - I(\epsilon_i^{**}\leqslant 0)\right\}\mathrm{d}s$$
$$= V_{1n}^{**}(\delta) + V_{2n}^{**}(\delta).$$

Similarly to the proof of Lemma A1, we obtain $\sup_t\left|\mathrm{pr}_{r|z}\{V_{1n}^{**}(\delta)\leqslant t\} - \mathrm{pr}_z\{-\delta^T H\leqslant t\}\right| = o_{\mathrm{pr}_z}(1)$.

Similarly to the proof of Lemma A2, we have $V_{2n}^{**}(\delta) = \delta^T B_1\delta/2 + o_{pr}^*(1)$. For n sufficiently large, on the event A_n, $\mathrm{sign}(\breve\beta_j^*) = \mathrm{sign}(\beta_{0j})$ and $\breve\beta_j^* = \bar\beta_{0j}$ for $j = 1,\ldots,q$; and $\breve\beta_j^* = 0$ for $j = q+1,\ldots,p$. Conditional on the data, $\lambda_n\sum_{j=1}^{p}\left\{|\breve\beta_j^* + \delta_j/n^{1/2}| - |\breve\beta_j^*|\right\} \to \lambda_0\sum_{j=1}^{p}\left\{|\delta_j|I(\breve\beta_j^* = 0) + \delta_j\,\mathrm{sign}(\beta_{0j})I(\breve\beta_j^* \neq 0)\right\}$. For any $1\leqslant j\leqslant p$,

$$\mathrm{pr}\left\{|\delta_j|I(\breve\beta_j^* = 0) + \delta_j\,\mathrm{sign}(\beta_{0j})I(\breve\beta_j^*\neq 0) = |\delta_j|I(\beta_{0j}=0) + \delta_j\,\mathrm{sign}(\beta_{0j})I(\beta_{0j}\neq 0)\right\}$$
$$\geqslant \mathrm{pr}\left\{|\delta_j|I(\breve\beta_j^* = 0) + \delta_j\,\mathrm{sign}(\beta_{0j})I(\breve\beta_j^*\neq 0) = |\delta_j|I(\beta_{0j}=0) + \delta_j\,\mathrm{sign}(\beta_{0j})I(\beta_{0j}\neq 0), A_n\right\} \to 1$$

as $n\to\infty$. Therefore, conditional on the data, as $n\to\infty$,

$$Q_n^{**}(\delta) \to -\delta^T H + \delta^T B_1\delta/2 + \lambda_0\sum_{j=1}^{p}\left\{|\delta_j|I(\beta_{0j}=0) + \delta_j\,\mathrm{sign}(\beta_{0j})I(\beta_{0j}\neq 0)\right\}$$

in distribution. Following the same argument as in Lemma 2 and applying epi-convergence theory, as in the unpublished technical reports of Geyer and Knight, the result is established by the equivalent representation of bootstrap consistency in (23.2) of van der Vaart (1998). □

References

Buchholz, A., Holländer, N. & Sauerbrei, W. (2008). On properties of predictors derived with a two-step bootstrap model averaging approach – A simulation study in the linear regression model. *Comp. Statist. Data Anal.* **52**, 2778–93.

Camponovo, L. (2015). On the validity of the pairs bootstrap for lasso estimators. *Biometrika* **102**, 981–7.

Chatterjee, A. & Lahiri, S. (2010). Asymptotic properties of the residual bootstrap for lasso estimators. *Proc. Am. Math. Soc.* **138**, 4497–509.

Chatterjee, A. & Lahiri, S. (2013). Rates of convergence of the adaptive lasso estimators to the oracle distribution and higher order refinements by the bootstrap. *Ann. Statist.* **41**, 1232–59.

Chatterjee, A. & Lahiri, S. N. (2011). Bootstrapping lasso estimators. *J. Am. Statist. Assoc.* **106**, 608–25.

Cheng, G. & Huang, J. Z. (2010). Bootstrap consistency for general semiparametric M-estimation. *Ann. Statist.* **38**, 2884–915.

De Bin, R., Janitza, S., Sauerbrei, W. & Boulesteix, A.-L. (2015). Subsampling versus bootstrapping in resampling-based model selection for multivariable regression. *Biometrics* **72**, 272–80.

Feng, X., He, X. & Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika* **98**, 995–9.

Ihorst, G., Frischer, T., Horak, F., Schumacher, M., Kopp, M., Forster, J., Mattes, J. & Kuehr, J. (2004). Long- and medium-term ozone effects on lung growth including a broad spectrum of exposure. *Eur. Respir. J.* **23**, 292–9.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755–70.

Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.

Koenker, R. (2016). *quantreg: Quantile regression*. R package version 5.35.

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

Koenker, R., Chernozhukov, V., He, X. & Peng, L., eds. (2017). *Handbook of Quantile Regression*. Boca Raton, Florida: Chapman & Hall/CRC.

Lee, E. R., Noh, H. & Park, B. U. (2014). Model selection via Bayesian information criterion for quantile regression models. *J. Am. Statist. Assoc.* **109**, 216–29.

Leeb, H. & Pötscher, B. M. (2008). Sparse estimators and the oracle property, or the return of Hodges' estimator. *J. Economet.* **142**, 201–11.

MINNIER, J., TIAN, L. & CAI, T. (2011). A perturbation method for inference on regularized regression estimates. *J. Am. Statist. Assoc.* **106**, 1371–82.
PÖTSCHER, B. M. & SCHNEIDER, U. (2009). On the distribution of the adaptive lasso estimator. *J. Statist. Plan. Infer.* **139**, 2775–90.
VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.
VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.
WU, Y. & LIU, Y. (2009). Variable selection in quantile regression. *Statist. Sinica* **19**, 801–17.
ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Statist. Assoc.* **101**, 1418–29.
ZOU, H. & YUAN, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36**, 1108–26.

[*Received on* 10 *August* 2017. *Editorial decision on* 24 *April* 2018]